Measuring Situation Assessment Performance through the Activities of Interest Score

John Salerno
Air Force Research Laboratory
Rome, New York, U.S.A.
John.Salerno@rl.af.mil

Abstract – ***Situation Awareness involves both the ability to identify and recognize the given activities and in assessing their importance through situation assessment. In this paper we look at a number of metrics that can be used in an information fusion framework to evaluate how well our assessment tools work at discriminating information from data. We begin our discussion by first providing a set of definitions and a reference model. We then summarize previous metrics that we have defined and describe how they can be applied to evaluating situation assessment capabilities. We conclude the paper with an example demonstrating their use and the meaning of the results.***

Keywords: Data Fusion, JDL, Endsley, Situation Awareness, Situation Assessment, Impact Assessment, Threat Assessment, Projection, Anticipation, Metrics

1 Background

Over the years, more than thirty fusion models have been proposed and countless research initiatives and personnel have attempted to define these models in great detail. Many of these models are tailored to the data, the fusion algorithms, and the context of the experiments. However, no model has become as influential in Data Fusion as the Joint Director's of Laboratories (JDL). The JDL model [1] has five levels: Level 0 – Sub-Object Data Assessment; Level 1 – Object Assessment; Level 2 – Situation Assessment; Level 3 – Impact Assessment; and Level 4 – Process Refinement.

A stream of data enters the model at level 0, Sub-Object Data Assessment. Level 0 provides physical access to the raw bits or signal. In addition, estimation and prediction of the existence of an object is performed based on pixel or signal level data association and characterization. Based on this low level data, objects are correlated and tagged over time in an attempt to build tracks and to perform object identification during level 1 processing, or Object Assessment. During Situation Assessment, or level 2 processing, knowledge about objects, their characteristics, their relationships with each other and their cross force relations are aggregated in an attempt to understand the current situation. Previously discovered or learned models generally drive this assessment. After Situation Assessment, the impact of the given situation must be assessed (Level 3 – Impact Assessment). The impact estimate can include likelihood estimates and cost/utility measures associated with the potential outcomes of a player's planned actions. It is important to note here that the JDL defined impact/threat assessment based on the perceived future. The final level, Process Refinement, provides a feedback mechanism to each of the other layers, including the sensor itself. It is important to note that transitions between layers and data input and output functions usually involve data reduction, either through metrics, filtering, or associations. While the JDL is a process model there are open questions on how to design a functional fusion system based on this guidance. Answers towards a functional fusion system will come with experimentation, evaluation, and validation of successful instances.

The JDL model can be considered as a data driven or bottom up model. The question becomes where does the context or meaning of the domain come that describes the relationships for level 2? How do we make sense of the data as it pertains to us? Endsley [2] defined a model that addresses Situation Awareness from this viewpoint (i.e., Mental Model). Her model has two main parts: the core Situation Awareness portion and the various factors affecting Situation Awareness. The core portion follows Endsley's proposition that Situation Awareness has three levels of mental processing: (1) Perception, (2) Comprehension, and (3) Projection. The second and much more elaborate part, describes in detail the various factors affecting Situation Awareness. Endsley defines *Situation Awareness* as a "state of knowledge that results from a process". This process, which may vary widely among individuals and contexts, is referred to as *Situation Assessment*, or as the process of achieving, acquiring, or maintaining Situation Awareness.

According to Endsley, Situation Awareness begins with perception. *Perception* provides information about the status, attributes and dynamics of the relevant elements in the environment. It also includes the classification of information into understood representations and provides the basic building blocks for comprehension and projection. Without a basic perception of important information, the odds of forming an incorrect picture of the situation increase dramatically. *Comprehension* of the situation encompasses how people combine, interpret, store, and retain information. Thus, it includes more than perceiving or attending to information; it includes the integration of multiple pieces of information and a determination of their relevance to an individual's underlying goals. Comprehension yields an organized

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE
JUL 2008

2. REPORT TYPE

3. DATES COVERED
00-00-2008 to 00-00-2008

4. TITLE AND SUBTITLE
Measuring Situation Assessment Performance through the Activities of Interest Score

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory,Rome,NY,13441

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR’S ACRONYM(S)

11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
11th International Conference on Information Fusion, June 30 ? July 3, 2008, Cologne, Germany.

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:

a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**8**	

picture of the current situation by determining the significance of objects and events. Furthermore, as a dynamic process, comprehension must combine new information with already existing knowledge to produce a composite picture of the situation as it evolves. Situation Awareness refers to the knowledge of the status and dynamics of the situational elements and the ability to make predictions based on that knowledge.

McGuinness and Foy [3] extended Endsley's Model by adding a fourth level, which they called *Resolution*. This level provides awareness of the best path to follow to achieve the desired outcome to the situation. Resolution results from drawing a single course of action from a subset of available actions. McGuinness and Foy believe that for any fusion system to be successful, it must be resilient and dynamic. It must also address the entire process from data acquisition to awareness, prediction and the ability to request elaboration or additional data and finishing with an appropriate action. McGuiness and Foy put Endsley's model and their model into perspective with an excellent analogy. They state that Perception is the attempt to answer the question "What are the current facts?" Comprehension asks "What is actually going on?" Projection asks "What is most likely to happen if…?" And Resolution asks "What exactly shall I do?"

Another point to be made is that any proposed model should not promote a serial process, but rather a parallel one. Neither the JDL Model nor Endsley suggest otherwise. Each function (for example in Endsley's model: Perception, Comprehension, Projection and Resolution) happens in parallel with continuous updates provided to and from each other. It should also be emphasized that each sub-component (in both models) also continuously interacts with each other and embarks its data/knowledge to the others.

Up until the break of the 21st century much activity was being spent in what we now call Sensor Fusion or by JDL's definition Level 0 and 1. Since then, emphasis has changed in an attempt to address data overload that exceeds the reasoning process of a user. The hope is that by combining data into more complex "objects" (groups, activities, or situations) one could significantly reduce data overload while increasing overall understanding – thus the increased attention in situation, impact and threat assessment.

In [7], we introduced the concept of the ***Data/Information Ratio*** in an attempt to measure this improvement. It is difficult to discuss situation or impact/threat assessment without some context. Recent work has concentrated on the Maritime and Cyber domains. Each of these domains has added to our understanding of the meaning of Situation Awareness.

In section 2 we present a number of definitions in hope to understand what Level 2 and 3 should entail. Based on these definitions we present a revised model and then summarize a set of metrics that have been developed (and used) to provide us an insight into how well the system is performing. We conclude this paper with an example demonstrating the use of the proposed metrics and discuss what insight they provide us about how well our system is performing.

2 Building the Combined Model

There continues to be a debate as to what Levels 1 and 2 represent. One belief is that Level 1 deals only with the tracking and identification of individual objects while Level 2 is the aggregation of the objects into groups or units. For example, Level 1 objects could be various equipments (tanks, APCs, missiles, etc). At Level 2, equipment along with personnel can be aggregated into a unit or division based on time and space. In our case, we believe a number of these products are created at Level 1 while others are Level 2. Level 1 attempts to answer such questions as Existence and Size Analysis (How many?), Identity Analysis (What/Who?), Kinematics Analysis (Where?), and When? But if we consider this separation then several questions arise; how do we account for concepts or non-physical objects and can't we track a group or activity like an object? What is a situation? How does the system acquire the necessary a priori knowledge (or relationships) to perform aggregation? What is the difference between models for identifying an object, a group or an activity?

To begin to answer these questions we first present a number of basic definitions and then use them to refine what we mean by Level 1 and 2. We then will explore the difference between Level 2/3 and what Endsley referred to as Projection. Section 3 will utilize these definitions in defining a set of metrics for evaluation.

In [4] an **entity** is defined as "something that has a distinct, separate existence, though it need not be a material existence. In particular, abstractions and legal fictions are usually regarded as entities. In general, there is also no presumption that an entity is animate. The word entity is often useful when one wants to refer to something that could be a human being, a non-human animal, a non-thinking life-form such as a plant or fungus, a lifeless object, or even a belief." An object is "a physical entity; something that is within the grasp of the senses" [4]; "something perceptible by one or more of the senses, especially by vision or touch" (The Free Dictionary). What if the entity is not a physical object? How can we describe it? Generally speaking, an abstract entity still can be associated with a time or existence and an abstract concept (e.g., a phone call, financial transaction, etc.)

A **group** is "a number of things being in some relation to each other". A group can be an interest group (terrorist cell, religious order), organizational group (police, government, Non-Governmental Organization (NGO) or military). An **event** is "something that takes place; an occurrence at an arbitrary point in time; something that happens at a given place and time" [4]. Both entities and groups can be associated with a specific event or events. Snidaro, Belluz and Froresti [5], further decompose an event into 3 classes: Simple, Spatial, and Transitive.

They define a *Simple event* as one which involves only a single entity with no interaction with others; a *Spatial event* describes events that occur in space and includes location. The third event, *Transitive*, involves two entities that are connected by some interaction. Spatial events can be a tank (entity) or unit (group) being at a given location at a specified time. If the tank or unit then interacts with another tank or unit then we say we have a Transitive event.

An **activity** is "something done as an action or a movement" [4]. Activities are composed of entities/groups related by one or more events over time and/or space. Thus, by definition an event, group or activity can be considered a complex entity (or in terms of the JDL, an object) and can be tracked and identified as such. As a side note, the JDL Lexicon defines an entity as "Any object or object set (or event or event set) which forms the basis of a hypothesis used in data fusion processes" but does not define what an object or event is. Now back to our discussion. By using the definitions presented above, we argue that activities and the aggregation of these activities (which we refer to as the situation) is both a part and a result of Level 1.

Models or a priori knowledge is necessary for level 1 to be capable of identifying the object, group or activity. This a priori knowledge (i.e., the relationships or associations) can be learned through Knowledge Discovery and validated by an operator or provided directly. Here we note that Knowledge Discovery techniques only learn statistically relevant occurrences. As such, new or novel ideas cannot be learned and require knowledge elicitation or conjecture of possible existence by a human. So what are examples of activities? Classical activities can range from a conventional war, potential multi-stage or coordinated cyber attacks, potential terrorist attacks (asymmetric) to operations other than war. These activities are composed of a number of interconnected and inter-related events. It should also be noted that the adversary is becoming more fluid with their Tactics, Techniques and Procedures (TTPs) and move from one classical activity to another (start with conventional and drop back to asymmetric, begin with cyber and move to asymmetric or a combination of one or more). No more will they use only one approach.

We define a **situation** as a *person's world view of a collection of activities that one is aware of at an instance in time.* We also argue that a computer system can identify an activity is occurring based on some a priori knowledge and interconnect a number of objects/events but cannot itself develop or provide Situation Awareness; only a person can be aware. A computer is a tool that can assist/support a person in developing awareness. Thus we argue that there is one situation or world view per person based on their context. **Shared Situation Awareness** is then a consensus view of a number of individual views on a specific activity or set of activities. Likewise, there is a growing community supporting "Shared action plans" to represent the group decision-making over the jointly observed information or data reduction.

The JDL Level 2 definition does not distinguish between time, current or future, while Level 3, Impact/Threat Assessment is specifically associated with the future. Why can't we have a current threat or impact? How is the current situation different from the projected or forecasted one? Can we have different impacts/threats depending on the timeframe that we are projecting? Based on experience we have found that it makes sense to split assessment based on time not functionality. This is similar to Endsley's comprehension and projection levels. We look at JDL Level 2 or Endsley's comprehension level as addressing the current situation (to include identification of potential impact/threat) and JDL Level 3, Endsley's projection level as the projection of the current situation and its analysis (i.e., future impacts and threats).

Bosse, Roy and Wark [6] define **Situation Assessment** as *"a quantitative evaluation of the situation that has to do with the notions of judgment, appraisal, and relevance."* Two products or components of situation assessment are: Impact and Threat assessment. Impact assessment is defined as *"the force of impression of one thing on another; an impelling or compelling effect. There is the notion of influence: one thing influencing another. In that sense, impact assessment estimates the effects on situations of planned or estimated/predicted actions by the participants, including interactions between action plans of multiple players."* They also define threat assessment as *"an expression of intention to inflict evil, injury, or damage. The focus of threat analysis is to assess the likelihood of truly hostile actions and, if they were to occur, projected possible outcomes…."* The only difference we note in their definition of impact/threat assessment is that again like the JDL definition, they are only concerned with the future.

Based on their definitions we can further define situation assessment as the understanding of the current situation, what it means to me (its impact/threat), the projection of the current situation into the future (which we refer to as the set of plausible futures) and the potential impacts/threats. In Level 2 or comprehension we need to have an understanding of "us" and what is important to "us" (commonly referred to in the literature as "Blue" but can also include "Grey"). In order to accomplish this we need to know such information as to our resources (capacity and capabilities), what is important to us (salience) and what our vulnerabilities are. Based on this information the identified activities within the current situation can be ranked based on their impact (associated damage) and threat (increased/decreased). This a priori knowledge gives us the insight necessary to assess how important each activity within the situation is and to what degree. These activities could then be ranked based on most likely (the impact) and most dangerous (the threat).

The next step would be to forecast or project the current situation into the future. To perform projection one would simply begin with the current situation and then generate from the a priori knowledge provided through the model a set of possible futures. Based on our understanding of "them" (commonly referred to as "Red" but can also include "Grey") we can then take this set of possible futures and generate the set of plausible ones. To acquire this subset we need to know "them" – their capacity/capability, intentions, and previous behavior. From this set of plausible futures we then perform a similar analysis as that was done with the current situation, identifying future impacts and threats. Both the projected impacts and threats on "us" and "them" are fed back to their respective capabilities such that projected increases/decreases are taken into account in future projections.

Thus, to summarize, a situation is a snapshot of the aggregated activities at time *t*. Projection takes the situation and projects or forecasts it to time *t* + *n*, where n is some number of time steps in the future. We would like to emphasize that the goal of generating plausible futures is not to provide the decision maker with the key to the future but to open their eyes to the possible and not to tunnel in on one or a subset of the possibilities. To this degree the decision maker can use this information to (1) develop a more robust set of actions, (2) ensure that their staff and personnel are trained for all possibilities and thus minimizing the chance of surprise and (3) better utilization of limited ISR assets. We also note that multiple, parallel sets and varying time increments can be generated and analyzed.

Feedback in any control system is very important, especially in an ever changing and dynamic environment. In this next section we discuss what type of feedback or what JDL calls Process Refinement (Level 4) means for Level 2/3 and conceptually how it can be implemented. We also present how it is affected by Projection.

The basic definition of **Process Refinement** as presented above covers two separate but integrated capabilities. For the purpose of our discussion we will divide them into external and internal process. Externally, we are concerned with providing sensors or collections with positioning information based on forecasted or anticipated movement of objects/entities or groups. The classical example here is the tracking of an object. A common tracking algorithm used in today's system is a Kalman Filter. Kalman Filters provide the ability to forecast where the object could be in one time increment in the future. This position information can then be provided to "better" position the sensor. Theoretically, a similar approach can be done with concepts and groups.

Recall our revised definition of Level 2 is concerned with assessment of the current situation. As one develops their understanding of the current situation, questions arise and more data could be required to either fill in the holes or reduce the uncertainty of the given data. These requirements can be considered as additional or revised *collection requirements* and be provided as feedback to the collection requirements process. Level 3 can provide similar data to the collection process, except from a somewhat different perspective. The projected activity or activities are just that. From a single current situation multiple futures can be developed. From each of these futures, the analyst can determine key events that could assist them in determining which one is unfolding. These key events can be used to drive the collection requirements process.

Internal processes also need to be monitored to ensure that the information processing system is performing as designed. At the object level one can suggest, possibly based on environmental inputs, which source is "better" at that time for tracking or identifying the object or sending the same sensor data to multiple algorithms (running in parallel), coming up with possibly different answers and combining the results in some manner. Similar concepts can be used at the activity level. As previously mentioned a second area is the update of *a priori* knowledge or models. As new information comes in and new knowledge is developed through the analysis and projection process, the analyst may update existing models or add/create new models (regardless of whether it is a new/modified object, group or activity). It is important to understand that tools can be provided (e.g., data mining, knowledge discovery, etc.) that can assist the user in finding new relationships or patterns but in many cases they also produce meaningless patterns or noise. In such a case it is important that the human use these tools as input and verify/validate the results. They are also limited to the data collected/available. Such tools cannot come up with "novel" or never before seen patterns or relationship where there is no data (or in most cases not statistically relevant) to support it. The human is still by far the most capable to develop such models and any technique/interface must take this into account. Figure 1 provides graphically how each of the components described above fit together.

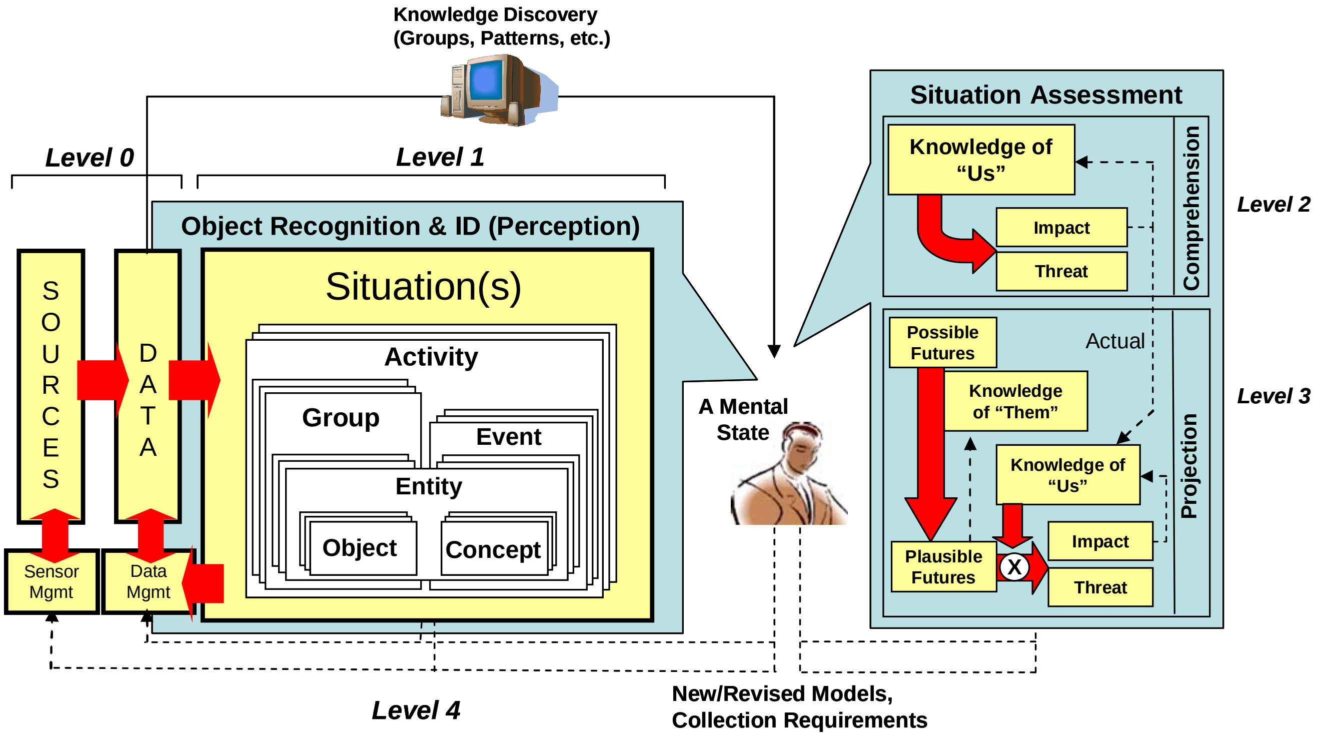

Figure 1. Revised Reference Model

3 Metrics

How well does a system work? Once we have defined the system and its purpose we can then develop metrics to evaluate how well it performs. Such metrics were described in [7] and classified into four categories or dimensions: (1) Confidence, (2) Purity, (3) Cost Utility and (4) Timeliness.

The first two measures tell us how well the lower level algorithms work in combining the observations into activities and thus how well our system is doing in correctly identifying and tracking the activities or overall situation. The third category provides us an indication as to how well the assessment capabilities work, while the fourth provides an overall measure of the performance of the system to provide the right information in a sufficient amount of time for the decision maker to make a decision.

Confidence is a measure of how well the system detects the true activities. It is composed of three factors: (1) Recall, (2) Precision and (3) Fragmentation. Recall measures the percentage of activities correctly identified by the Level 1 system (Correct Detections) in relation to the number of "real" activities as defined in the ground truth (Known Activities) and is defined in Equation 1.

$$\text{Recall} = \frac{\text{Correct Detections}}{\text{Known Activities}} \qquad (1)$$

Precision is the percentage of activities correctly identified by the Level 1 system (Correct Detections) in relation to the total number activities (Detected Activities) provided by the Level 1 system and is defined in Equation 2. Fragmentation is defined as the percentage of activities reported as multiple activities that should have been reported as a single activity and is defined in Equation 3. Confidence is typically reported as a probability.

$$\text{Precision} = \frac{\text{Correct Detections}}{\text{Detected Activities}} \qquad (2)$$

and

$$\text{Fragmentation} = \frac{\text{No. Results that Identify the same Known Activity}}{\text{Detected Activities}} \qquad (3)$$

The second dimension, *Purity* characterizes the quality of the proposed activities (i.e., were the observations correctly matched and connected to the right activity track). To measure purity we define a "*Mis-assignment Rate*" as the percentage of evidence or observations that were incorrectly assigned to a given activity and *Evidence Recall* as the percentage of evidence or alerts detected in relation to the "total known" events or observations.

The third dimension, *Cost Utility* was not used as defined in [7]. After implementing this metric, we found that it provided little information about the performance of a system and subsequently we redefined it under our cyber work as the *"Attack" score.* What we were trying to measure was how effective the system is in ranking the activities of interest in a predetermined order. For example, in a cyber domain we are interested in those activities (or what we called potential attacks) that were most detrimental to our operations based on either impact/threat or mission. The cyber work is described in detail in [9]. This ranking can be based on “most likely” or “most dangerous”. Based on this, we have again renamed this metric to “Activities of Interest (AOI)” Score which we believe more accurately describes what insight it provides and implies that we are not interested in attacks only.

The final and fourth dimension is Timeliness. This dimension attempts to measure the ability of the system to respond within time requirements of a particular domain. For example, we are interested in how quickly the system can identify the activity or activities of interest. But this value alone is meaningless unless we take into account whether this time is soon enough for us to take action. Combined it is hoped that these measures can correctly characterize how well the system performs. Figure 2 shows how the given metrics are mapped against our reference model.

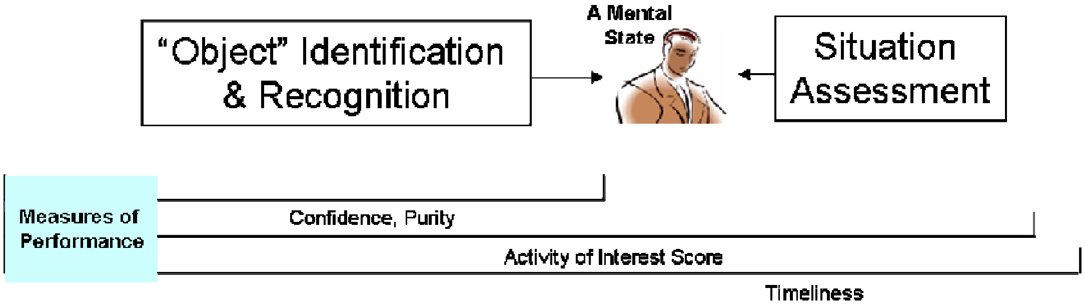

Figure 2. Metrics Mapped to Reference Model

3.1 Activities of Interest (AOI) Score

Focusing on the third category, how can the AOI score provide us insight into our assessment process? Recall that the objective of the AOI Score is one to measure how well the system has identified the activities that are most important to us. The criteria for what is most important are dependent on the decision maker’s or analyst’s desire. In determining potential adversary Courses of Actions (COAs) the decision maker is interested in which COAs are most likely and which are most dangerous – each answer using different information as part of the analysis and providing the decision maker with a different perspective. For example ranking the “most likely”, needs to take into account the adversary’s capability/capacity (what can they do), intent (what is/are their goals) and opportunity. Most dangerous must not only take into account the information required to determine the most likely but also Blue’s vulnerabilities and an assessment as to whether the adversary will be successful and to what extent will Blue incur damage. The ranking of each can also change between the current analysis and the projection of future impact/threat.

We will illustrate the basic concept through an example from our cyber work. In our example individual observations are provided by such systems as Intrusion Detection Systems (IDSs), Netflow, System Logs, etc. and aggregated into AOI tracks. This would be the output of our level 1 capability and can be simply displayed as a list (one form of output) of the activities. If no other knowledge is available, this list would be just that, a list of activities in no order with no indication of “importance”, however, in the real world, the number of activities can be very large (hundreds to thousands based on pruning or lack of) – some ordering is needed to place the most important ones that the analyst or decision maker should be concerned with at the top. In our cyber example we would like to prioritize the activities based on their current of future impact/threat. So how can we measure both what the additional knowledge provides and how well our ranking algorithms work?

3.2 Measuring How Well We Are Doing

We can compute the AOI score by using Equation 4 shown below.

$$\text{AOI Score} = \frac{\text{NAOI} * \text{NA} - \sum_{i=1}^{\text{NAOIR}} P_i}{\text{NAOI} * \text{NA} - \sum_{i=1}^{\text{NAOI}} i} \qquad (4)$$

where,

NAOI	Number of Activities of Interest in Ground Truth
NAOIR	Number of AOIs in Results
NA	Total Number of Activities in Ground Truth
Pi	Position of the i^{th} Activity of Interest

If any (or all) of the AOIs are not part of the results list or if their position is greater than the total number of activities in the ground truth, we set the position value (P_i) for those AOI(s) equal to the Total Number of Activities in the Ground Truth. By adding this condition, if there are no AOIs included/identified in the results list, the AOI score will equal 0. Whereas, if there is only a subset of the actual AOIs identified, the system will get credit for only those.

Let us consider a simple example. Given the ground truth and the results provided as output by our Level 1 system (Situation Recognition and Identification) without any further knowledge or assessment as shown in Table 1, we can compute a baseline. This baseline can then be used to determine the value added by additional knowledge and processing provided by situation assessment.

We first compute how well our level 1 system has identified the activities that make up what we call the situation. To compute recall we use Equation 1 (6/8 = 0.75 or 75%). In this case, our system has identified more activities than what actually is occurring (one not even there, a false positive and the other is a fragment). Optimum recall would be 100%. Let us next examine *precision*. We compute precision using Equation 2 (6/6=1.0 or 100%). In this case the system has correctly identified all the activities in the ground truth. The last metric we will examine within the first dimension is fragmentation. *Fragmentation* is computed using Equation 3 (1/8= .125 or 12.5%). In summary if we had a capability that provided the results as shown above we would have a recall of 75%, a precision of 100% and a fragmentation of 12.5%. What do these numbers tell us? There were 25% more activities identified than there actually were and of the 25% half of them (12.5%) were because they should have been associated with an existing activity (fragments). Let us now consider the question "Of the activities that have been identified are there any of them that we should be concerned with?" Our next metric, AOI score, will tell us how we are doing in answering this question.

Table 1. Data for Example

Ground Truth

GT0	Activity 1 (AOI 1)
GT1	Activity 2 (AOI 2)
GT2	Activity 3
GT3	Activity 4
GT4	Activity 5
GT5	Activity 6

Proposed Activities

R0	Activity 4
R1	Activity 3
R2	Activity 2 (AOI 2)
R3	Activity not part of GT
R4	Activity 5
R5	Activity 1 (AOI 1)
R6	Activity (Frag, should be part of Activity 2)
R7	Activity 6

Let us first compute a baseline AOI score where we have no further information about the activities. To compute the AOI score we have the number of activities of interest in the ground truth as 2 and the total number of activities in the ground truth as 6. The remaining 4, that we do not consider to be AOIs, are either activities that have no or minimal impact or threat to us. We next need to know the ordering the system identifies for those activities that are of interest (as identified by in the results). In our example the first "important" activity is third in the proposed list while the second "important" activity is in the sixth position (Table 1). We simply add these two values together for a value of 9. The final value we need to compute is just the geometric sum of the "important" activities which is simply 2 activities: 2+1=3. Substituting in the values into Equation 4 we have a value of ((2)(6)-9)/((2)(6)-3) = 3/9 or 0.33. This means that an analyst or user would have to consider roughly 2/3'rds of the activities before "seeing" the most important activities.

Now assume that we are provided additional knowledge about the activities of Table 1 after further analysis is provided by Level 2, Situation Analysis. For example if we were looking at a list of cyber activities we might have more information on the computers (their connectivity, operating system, services and applications) and what mission they perform. Using this knowledge we can then determine how important the activity is based on impact or threat to the mission. Vulnerabilities come in to play also; for instance, am I vulnerable to the attack being executed against me making that activity more important if I am vulnerable and less important if not. To compute the value of this additional information in the ranking we simply apply Equation 4, a second time. Table 2 provides an example.

Table 2. Data for Example 2

Proposed Activities

R0	Activity 4
R1	Activity 2 (AOI 2)
R2	Activity 1 (AOI 1)
R3	Activity 3
R4	Activity not part of GT
R5	Activity 5
R6	Activity (Frag, should be part of Activity 2)
R7	Activity 6

The only value that changes in our equation is the sum of the positions of the activities of interest. This new value is 2+3=5 and our new value for the AOI score becomes ((2)(6)-5)/((2)(6)-3) = 7/9 or 0.78. One can easily see that if we got the most important activities in positions 1 and 2 then the AOI score would be 1.0. Similar computations can be made for Level 3, plausible futures and their impact/threat. Similar to Level 2 we need only ground truth for the given scenario. We note here that the AOI score does not distinguish between how important the AOIs are just the fact that they are AOI.

Extensions to this metric can be made to take into such a scenario.

4 Conclusion

The purpose of this paper was to present a view, based on a number of years of building systems which describes a more logical boundary for Levels 2 and 3 of the JDL Data Fusion model. The model has served the fusion community well but as time and understanding has progressed, so must the model. We began our discussion with a number of basic definitions, described some of the foreseen deficiencies with the existing model and based on these deficiencies presented a slightly altered/enhanced view capitalizing on the great work accomplished by both the JDL (for the current situation) and Dr. Endsley. The model as described provides the community with a process that is domain agnostic. It has been successfully used in cyber defense and is currently being applied to the asymmetric threat. We concluded this paper with a discussion on metrics and how one in particular, the AOI Score, can provide us with insight about the value added by the additional knowledge derived through situation assessment.

Acknowledgements

The author would like to thank Dr. Erik Blasch, AFRL/RYAA, Mr. George Tadda and Mr. Mike Hinman, AFRL/RIEA, Mr. Joe Antonik, AFRL/RISB, Dr Shanchieh Yang, RIT and Dr Moises Sudit and Mr. Adam Stotz, University of Buffalo for there reviews and comments.

References

[1] U.S. Department of Defense, Data Fusion Subpanel of the Joint Directors of Laboratories, Technical Panel for C3, "Data fusion lexicon," 1991.

[2] Mica R. Endsley. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors Journal*, Volume 37(1), pgs 32-64, March 1995.

[3] B. McGuinness, & J. L. Foy, *A subjective measure of SA: The Crew Awareness Rating Scale (CARS)*Proceedings of the first human performance, situation awareness, and automation conference, Savannah, Georgia, October 2000.

[4] http:\\www.wikipedia.org

[5] L. Snidaro, M. Belluz, G. Foresti, “Domain knowledge for security applications”, ISIF’07.

[6] E. Bosse, J. Roy, and S. Wark, “Concepts, Models, and Tools for Information Fusion”, Artech House, Inc., 2007, ISBN-13: 978-1-59693-081-0, pg 43.

[7] John J Salerno, Michael Hinman and Douglas Boulware, “Evaluating Algorithmic Techniques in Supporting Situation Awareness”, In *Proc of the Defense and Security Conference*, Orlando, FL, March 2005.

[8] E. Townsend, RISKS: The Key to Combat Intelligence, The Military Service Publishing Company, Harrisburg, Pennsylvania, 1955.

[9] Llinas, J. and Waltz, E.L (1990), "Multi-Sensor Data Fusion", Artech House, Norwood, MA, 1990.

[10] George Tadda, “Measuring the Performance of Cyber Situation Awareness Systems”, In *Proc of the Fusion Conference*, Cologne GE, 2008.

[11] Moises Sudit, Adam Stotz, and Michael Holender, William Tagliaferri, and Kathie Canarelli, “Situational awareness of a coordinated cyber attack”, In *Proc of the Defense and Security Conference (*5812, 114), Orlando, FL, March 2005.

[12] Jared Holsopple, Shanchieh Jay Yang, and Moises Sudit, “TANDI: threat assessment of network data and information”, In *Proc of the Defense and Security Conference(* 6242, 62420) Orlando, FL, March 2006.